3.3 WP 1.3 Data combination strategies: calibration, sampling, quality control, error analysis, …

Start date or starting event:	**KO**					
Duration	**0-36**					
Participant codes :		**P2**	**P3**	**P4**	**P5**	**P6**

Responsible: Partner 6 (Dipl. Met. M. Kästner, DLR-DFD)

3.3.1 Objectives

- Combination of data sets in different temporal or spatial resolution sometimes describing different physical observables usually leads to a loss of information or an enormous increase in data volume. Since all subsequent tasks like calibration, sampling, or error analysis build on such a structure, finding an agreement is essential. It allows an equivalent evaluation of all formats and their direct intercomparison with independent data sets.
- The choice of appropriate examination techniques should not be restricted to the partners but is a subject of discussion with the steering committee as well.

3.3.2 Methodology and scientific achievements

The combination of data sets at different temporal and spatial resolution in common areas, common periods and common grids are essential for an effective inter-comparison of sensors, of different rainfall algorithms and for validation with independent data. **Partner 6** coordinated a joint validation effort with rainfall estimations from **Partners 1, 2, 3, 4, 5 and 7**. A common database of case studies for validation and inter-comparison purposes wasestablished characterized by continuous statistics and categorical statistical measures. Some results are discussed in section 3.9.2

Various climate models predict a decrease of precipitation for future changes over many parts of the subtropics, particularly in the winter (Bolle (ed.) 2003). Therefore, it is essential to have not only climatological data over land but also over the oceans. Based on a quasi operational data production with the PR-Adjusted TMI Estimation of Rainfall (PATER) algorithm that combines TMI with PR TRMM data, a series of data sets (Aug. 2002 - June 2003) was processed by **Partner 6** with special interest on the Eastern North Atlantic and the Mediterranean (Fig. 3.7) in a high spatial resolution of ~28 km. The daily course of precipitation is not easily obtained from TRMM data because of the low orbit, but daily, pentad and monthly means of gridded PATER rainfall rates are re-sampled into a common grid for inter-comparison with other techniques.

The geographical distribution of monthly means of precipitation over water excluding the coastlines is shown for a) November 2002, b) January 2003, c) March 2003 and d) May 2003. During the winter rainy season the rain intensity steadily increases from October to March and then quickly drops to the lower May values. Note, that the rain rate is in mm/(8 h) due to three TRMM orbits per day and not in mm h^{-1} as indicated, therefore the maximum value of 1.2 on the linear scale corresponds to 110 mm precipitation per month.

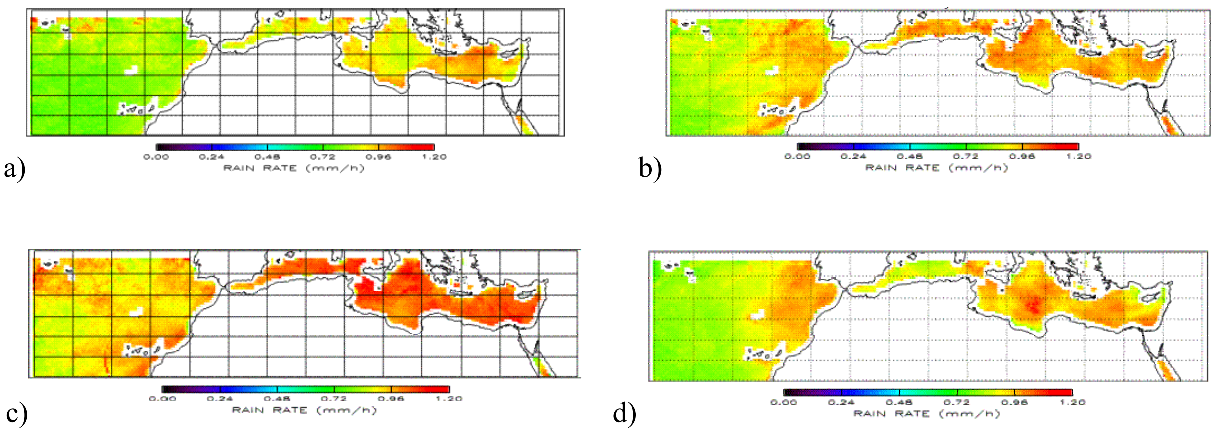

Figure 3.7. PATER retrieval. Geographical distribution of precipitation at 0.2 degree spatial resolution in the winter season 2002/2003, a) Nov. 2002, b) Jan. 2003, c) Mar. 2003, d) May 2003.

These retrieved precipitation data are compared to the standard TRMM 3B43 product. Fig. 3.8 shows these data at a 1 degree resolution for the same months. In this dataset the maximum rainfall is reached already in January and, consequently, the January precipitation distribution differs much more from that of May than in the case of PATER. PATER data over the Atlantic show an obvious E-W gradient to the coast of Africa, especially in January and May, which the 3B43 data do not reflect. On the other hand, both data sets, PATER and 3B43, show the gradient to the north over the Atlantic. The rain intensities are comparable: note the different colour code and maximum values.

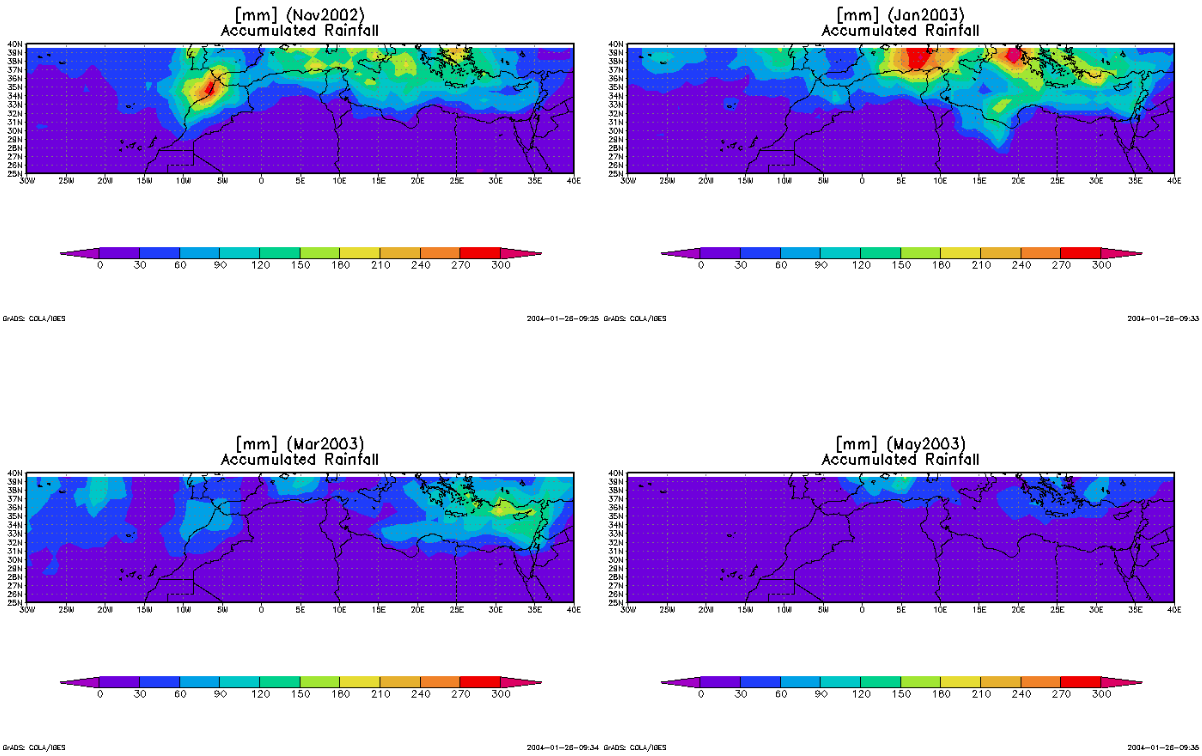

Figure 3.8. TRMM 3B43 monthly means in 1 degree spatial resolution

Since January 2002 experimental 3-hourly gridded, 0.25° lat-lon grid (~28 km) within the zonal bands of 60 N/S TRMM level 3 data are available. **Partner 6** collected the calibrated IR (3B41RT), combined MW (3B40RT) based on GPROF for SSM/I and TMI, and the merged IRMicro (3B42RT). Quasi-operational accumulated rainfall rates over Europe for 5 days (pentad) are produced showing the areas of high flood potential, although the precipitation is only one factor among others for a flood event. Fig. 3.9 gives the accumulated rain in mm for a pentad basing on TRMM 3B42RT data. Heavy rainfall led to locally flooding of the town Laudun-l'Ardoise (Fig. 3.10) located between Orange and Avignon at the Rhone river in southern France.

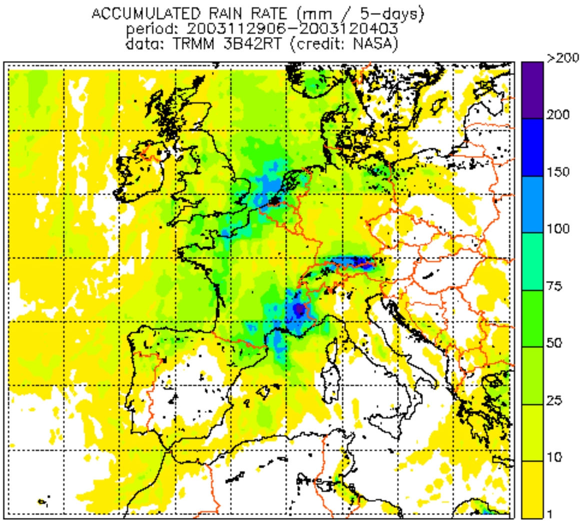

Figure 3.9. Rain rate accumulated over the pentad from 29 Nov 2003, 0600 UTC till 04 Dec 2003, 0300 UTC

Figure 3.10. Flooding of Laudun-l'Ardoise (F) on 4 Dec 2003 (credit: DLR/IKONOS)

3.3.3 Discussion and conclusion

The correlation coefficient r is often used in comparison tasks. It has already been shown that the inter-comparison of different rainfall algorithms are dependent on the grid size, here 0.25°. For the Algerian severe weather event the maximum r is reached for the both PMW techniques, namely r = 0.88 for PATER vs. the frequency difference algorithm (FDA). In this case the accuracy measured in terms of Heidke skill score (HSS), which is a combination of hits, false alarms, misses and correct negatives, is truly dependent on the threshold RR-min, discriminating rain from no-rain pixel. Fig 3.11 stems from varying the minimum detectable rain rate RR-min for the rain algorithm pair PATER vs. FDA. The event occurrences for each category are given on the y-axis. For RR-min = 1 mm h^{-1} false alarms and misses are of the same order, which means that there both the algorithms are unbiased. An important impact of RR-min to the accuracy is shown by the increase of HSS from 0.27 to 0.75 which means that higher rain rates are in better agreement than the lower ones.

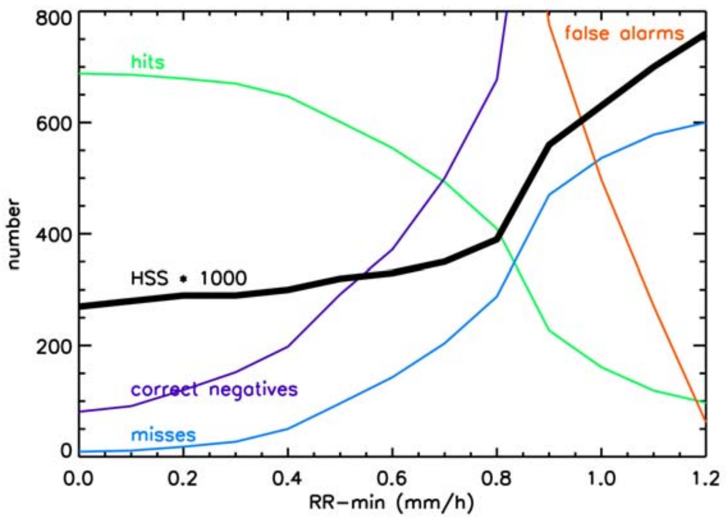

Figure 3.11. Comparison PATER vs. FDA. The accuracy via HSS*1000 (thick) is dependent on the rain/no-rain threshold RR-min.